Renewable and Sustainable Energy Reviews 31 (2014) 509–519

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models

Amit Kumar Yadav [a], Hasmat Malik [b], S.S. Chandel [a,*]

[a] Centre for Energy and Environment, National Institute of Technology, Hamirpur, 177005 Himachal Pradesh, India
[b] Department of Electrical Engineering, Indian Institute of Technology Delhi 110016, India

ARTICLE INFO

Article history:
Received 9 August 2013
Received in revised form
22 November 2013
Accepted 18 December 2013
Available online 20 January 2014

Keywords:
Solar energy
Solar potential
Solar radiation prediction
Artificial neural network
WEKA

ABSTRACT

The prediction of solar radiation is important for several applications in renewable energy research. Solar radiation is predicted by a number of solar radiation models both conventional and Artificial Neural Network (ANN) based models. There are a number of meteorological and geographical variables which affect solar radiation prediction, so identification of suitable variables for accurate solar radiation prediction is an important research area. With this main objective, Waikato Environment for Knowledge Analysis (WEKA) software is applied to 26 Indian locations having different climatic conditions to find most influencing input parameters for solar radiation prediction in ANN models. The input parameters identified are latitude, longitude, temperature, maximum temperature, minimum temperature, altitude and sunshine hours for different cities of India. In order to check the prediction accuracy using the identified parameters, three Artificial Neural Network (ANN) models are developed (ANN-1, ANN-2 and ANN-3). The maximum MAPE for ANN-1, ANN-2 and ANN-3 models are found to be 20.12%, 6.89% and 9.04% respectively, showing 13.23% improved prediction accuracy of the ANN-2 model which utilizes temperature, maximum temperature, minimum temperature, height above sea level and sunshine hours as input variables in comparison to the ANN-1 model. The WEKA identifies temperature, maximum temperature, minimum temperature, altitude and sunshine hours as the most relevant input variables and latitude, longitude as the least influencing variables in solar radiation prediction. The methodology is also used to identify the solar energy potential of Western Himalayan state of Himachal Pradesh, India. The results show good solar potential with yearly solar radiation variation as 3.59–5.38 kWh/m^2/day for a large number of solar applications including solar power generation in this region.

Contents

1. Introduction . . . 509
2. Literature survey for identification of input parameters for ANN based solar radiation prediction . . . 510
3. Methodology . . . 512
   3.1. Source of solar radiation data . . . 512
   3.2. Input variables selection using WEKA . . . 512
   3.3. Solar radiation prediction models with selected inputs . . . 513
4. Results and discussion . . . 515
   4.1. Solar radiation predicted by the ANN-3 model . . . 517
5. Conclusion . . . 517
Appendix A. MATLAB code for solar radiation using nftool . . . 518
References . . . 518

1. Introduction

Solar energy is a clean resource which has a vast potential to meet the energy needs. Solar potential assessment of a region requires information about the measured solar radiation at different locations. The solar radiation components are measured

* Corresponding author. Tel.: +91 1972 254748; fax: +91 1972 223834.
E-mail address: sschandel2013@gmail.com (S.S. Chandel).

1364-0321/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rser.2013.12.008

Nomenclature

BP	back propagation
H	altitude
Lat	latitude
Long	longitude
LM	Levenberg–Marquardt
MAPE	mean absolute percentage error
MBE	mean bias error
n	number of data samples
R	correlation coefficient
RMSE	root mean square error
SH	sunshine hours
$SR_{i(ANN)}$	predicted monthly average daily solar radiation data for month i.
$SR_{i(actual)}$	measured monthly average daily solar radiation data for month i
T	temperature
T_{max}	maximum temperature
T_{min}	minimum temperature

generally using pyranometer, solarimeter, pyroheliometer, etc. with the data acquisition system. It is not possible to install measuring instruments at every site due to high costs resulting in non-availability of measured solar radiation data for most sites worldwide. In addition, missing records in data set have been found due to lack of accuracy of measuring equipments. As such the solar radiation measured at nearby meteorological stations are used for solar radiation prediction, system design and installation. Solar radiation is predicted by a number of solar radiation models both conventional and Artificial Neural Network (ANN) based models. The prediction of solar radiation has generated a renewed interest in recent years, mostly due to its relevance in renewable energy research and applications. Bakirci [1] reviewed several empirical models in terms of cloudiness, evaporation, total precipitation, latitude, altitude, number of rainy days, relative humidity, soil temperature, maximum and mean temperature, sunshine hours and extraterrestrial radiation for estimating solar radiation. Due to lack of accuracy of empirical models, ANN techniques have been used for prediction as these give better results than empirical models [2–6]. There are a large number of meteorological and geographical variables which affect solar radiation and various researchers have used different variables for solar radiation prediction. In a comprehensive review for the solar radiation prediction using ANN techniques Yadav and Chandel [7] have pointed out that there is a need to identify the most influencing parameters which are to be used for solar radiation prediction. In order to find the most relevant input parameters, one has to select variables by combining different input parameters that provide best prediction which is time consuming. Therefore, in this study Waikato Environment for Knowledge Analysis (WEKA) software version 3.7.10 is used to identify the most relevant input parameters for solar radiation prediction of 26 Indian locations with different climatic conditions as a follow up of our study. However, this methodology can be used for other locations worldwide.

In order to check the prediction accuracy, three ANN models are developed (ANN-1, ANN-2 and ANN-3). The ANN-1 model is developed using all input variables. The ANN-2 model is developed using most relevant input parameters given by WEKA. The ANN-3 model is developed neglecting sunshine hours from relevant input parameters so that it can be used to predict solar radiation at Indian sites where sunshine database are not available. In addition, the ANN-3 model is used to predict solar radiation in mountainous region of Himachal Pradesh (30.38–33.21° N Lat and 75.77–79.07° E Long), India for identifying solar energy resource potential, as no study has been reported so far except [8,9], which uses National Aeronautics and Space Administration (NASA) values for solar energy potential. The NASA compiled solar radiation data are easily available, but there is a lack of accuracy due to indirect interpretation of data observed from space and calculation from snapshot images (pixel wise) [10]. The root mean square error between NASA solar data and measured solar radiation data for Indian cities is found to vary for Indian locations from 0.177 to 0.416 kWh/m^2 as shown by Karakoti et al. [11].

This paper is organized as follows: the literature review is given in Section 2.The database and methodology used are presented in Section 3. The results are presented and discussed in Section 4 and conclusion in Section 5.

2. Literature survey for identification of input parameters for ANN based solar radiation prediction

The ANN models use different meteorological and geographical variables of a location as inputs for the prediction of solar radiation as discussed in Ref. [7]. AI-Alawi and AI-Hinai [12] discussed multilayer feed forward network, back propagation (BP) training algorithm for global radiation prediction in Seeb. The inputs used in network are location, month, mean pressure, mean temperature, mean vapor pressure, mean relative humidity, mean wind speed and mean sunshine hours. The MAPE varies from 5.43 to 7.30. Sözen et al. [13,14] used meteorological and geographical data as input variables in the ANN model for solar radiation estimation in Turkey. The transfer function for model is logistic sigmoid and learning algorithm is Scaled conjugate gradient, Pola-Ribiere conjugate gradient Levenberg–Marquardt. The MAPE value for MLP network is 6.73%.

Mohandes et al. [15] applied ANN for global solar radiation modeling in Saudi Arabia as a function of latitude, longitude, altitude and sunshine duration. The results show that network with 4,10,1 neurons in input, hidden, output layers perform best and in testing stations MAPE changes from 6.5 to 19.1. Ouammi et al. [16] used ANN for estimation of solar radiation for 41 Moroccan sites. The network utilized input parameters as normalized values of longitude, latitude and elevation. The predicted solar irradiation varies from 5030 to 6230 Wh/m^2/day.

Rehman and Mohandes [17] applied four combinations of input parameters: day, maximum air temperature, mean air temperature, relative humidity, for estimating diffuse solar radiation for Abha city in Saudi Arabia. It is discovered that using relative humidity and daily mean temperature gives better results than other combinations with mean square error (MSE) of 5.18×10^{-7}.

Azeez [18] employed Feed forward back propagation Neural Network for monthly estimation of average global solar irradiation in Gusau, Nigeria. The sunshine duration, maximum ambient temperature and relative humidity are taken as input and solar radiation as output. The statistical analysis (R=99.96, MPE=0.8512, RMSE=0.0028) has shown good agreement between the estimated and measured values of global solar radiation.

Linares-Rodriguez et al. [19] used the MLP model for estimating solar radiation over Spain from satellite-obtained irradiances. The input layer has 12 inputs (11 Meteosat channels and clear sky solar

Table 1
Input variables used in ANN based prediction of solar radiation.

Reference	Models and training algorithm	Input variables to ANN Model	ANN Model prediction accuracy	Location
Linares-Rodríguez et al. [22]	MLP	Latitude, longitude, day of the year, daily clear sky global radiation, cloud cover, total column ozone and water vapor	RMSE 13.52% for training stations and 14.20% for testing stations	Spain
Koca et al. [23]	MLP	Latitude, longitude, altitude, months, average temperature, average cloudiness, average wind velocity and sunshine duration	Maximum RMSE is 6.9%	Seven cities in Mediterranean region of Anatolia, Turkey
Khatib et al. [24]	MLP	Latitude, longitude, day's number and sunshine ratio	The MAPE in estimating global and diffuse radiation are 7.96%, 9.8% respectively	Malaysia
Khatib et al. [25]	Linear, nonlinear, fuzzy logic and ANN models	Latitude, longitude, day number and sunshine ratio	MAPE of 5.38% (global radiation), 1.53% (diffuse radiation)	Five sites in Malaysia
Elminir et al. [26]	Multilayer feed forward network	Wind direction, wind velocity, ambient temperature, relative humidity, cloudiness and water vapor	RMSE are 5.02%, 7.46% and 3.97% for infrared, ultraviolet and global solar radiation respectively	Helwan, Aswan monitoring stations
Tymvios et al. [27]	ANN and Angström	Theoretical daily sunshine duration, measured daily sunshine duration, month, daily maximum temperature, monthly mean value of theoretical sunshine duration, monthly mean value of measured sunshine duration, extraterrestrial radiation, monthly mean value of daily global radiation, total global radiation, daily extraterrestrial radiation.	The maximum RMSE of ANN model is 10.15 and in Ångström Model, RMSE is 13.36, showing ANN model give better results than Ångström Model	Athalassa in Cyprus
Alam et al. [28]	MLP	Latitude, longitude, altitude, month of the year, mean duration of sunshine per hour, rainfall ratio, relative humidity	RMSE varies from 1.65 to 2.79%	India
Jiang [29]	Feed-forward back propagation neural network and Empirical Model	Monthly mean daily clearness index, sunshine percentage	RMSE in empirical models are 0.783, 0.781 whereas in ANN model is 0.746, showing accurate estimation of ANN than empirical models.	China
Mubiru and Banda [30]	Feed forward back-propagation ANN; Levenberg–Marquardt (LM)	Annual average of sunshine hours, cloud cover, relative humidity, rainfall, latitude, longitude and altitude	MAPE, R^2 are 0.3, 97.4% respectively and better results obtained by ANN than sunshine based conventional model	Uganda
Şenkal and Kuleli [31]	ANN and Physical model	Latitude, longitude, altitude, month, mean diffuse radiation and mean beam radiation	RMSE values using the MLP and the physical model are 54 W/m^2 and 64 W/m^2 (training cities); 91 W/m^2 and 125 W/m^2 (testing cities), respectively	Turkey
Jiang Y [32]	ANN model and empirical regression model	Latitude, altitude and mean sunshine	$R^2=0.97$, RMSE = 1.4 MJ/m^2	China
Benghanem et al. [33]	ANN model	Different combination of air temperature, relative humidity, sunshine duration and the day of year	R value of 97.65% is obtained using sunshine duration and air temperature as inputs to the ANN model	Al-Madinah (Saudi Arabia)
Fadare [34]	ANN	Latitude, longitude, altitude, month, mean sunshine duration, mean temperature, and relative humidity	The R^2 for training and testing cities are higher than 90%	195 cities in Nigeria
Azadeh et al. [35]	Integrated ANN-MLP	Location, month, mean value of maximum temperature, minimum temperature, relative humidity, vapor pressure, total precipitation, wind speed and sunshine hours	MAPE is 0.03 and ANN models give better results than Angstrom model	Iran
Sözen and Arcak-lioğlu [36]	ANN	Geographical coordinates, mean sunshine duration, mean temperature and month	MAPE is less than 3.832%,	Turkey
Rahimikhoob [37]	ANN	Maximum and minimum air temperature, extraterrestrial radiation	RMSE 2.534 MJ/m^2/day, R^2 88.9% better than Hargreaves and Samani [38] equation	Ahwaz (Iran)
Hasni et al. [39]	ANN	Air temperature, relative humidity	The MAPE, R^2 are 2.9971%, 99.99%	south-western region of Algeria
Yildiz et al. [40]	Two ANN models	Latitude, longitude, altitude, month and meteorological land surface temperature to first ANN Model. latitude, longitude, altitude, month and satellite land surface temperature to second ANN Model.	The R^2 for first, second ANN are 80.41%, 82.37% respectively for testing station, showing better estimation of second model than first model	Turkey
Rumbayan et al. [41]	MLP	Month, latitude, wind speed, precipitation, sunshine duration, humidity and temperature	MAPE is found to be 3.4% with 9 neurons in hidden layer	Indonesia

radiation). The RMSE is 6.74%. The model performs well in cloudy and clear sky condition.

Kisi [20] investigated fuzzy genetic for solar radiation modeling of seven cities in Turkey. The authors selected latitude, longitude, altitude, month as inputs and RMSE is 6.29 MJ/m^2. It is shown that the fuzzy genetic method gives better results than the ANN and ANFIS (adaptive neuro fuzzy inference system) model.

Mostafavi et al. [21] used hybrid genetic programming (GP) and simulated annealing (SA) called as GP/SA for new formulation of solar radiation in terms of sunshine, total precipitation, mean relative humidity, maximum and minimum temperature. The MAPE varies from 0.103 to 0.214 for Tehran and Kerman cities in Iran. It is suggested that maximum and minimum temperature are most influencing variable in prediction. The different ANN inputs parameters and prediction accuracy are summarized in Table 1.

Based on the literature survey it is found that prediction accuracy of ANN models get changed with geographical and meteorological variables as input parameters. For selection of

Fig. 1. Map of India showing selected cities for training and testing the ANN model.

relevant input parameters the researcher has to use different combinations of ANN input parameters to evaluate prediction accuracy of ANN models which requires large computational analysis. Therefore, the selection of most relevant input parameters for ANN models is an important research gap which is undertaken in the present study.

3. Methodology

3.1. Source of solar radiation data

The 26 selected cities located in different climatic zones of India are used for training and testing in ANN models as shown in Fig. 1.

The temperature data (T (°C), T_{max} (°C) and T_{min} (°C)) of these stations are taken from National Aeronautics and Space Administration (NASA) [42]. The sunshine hour and monthly average daily solar radiation data (kWh/m^2/day) are taken from Centre for Energy and Environment, National Institute of Technology, Hamirpur H.P. India, Indian Meteorological Department (IMD) Pune compiled by Anna Mani [43] and solar radiation handbook [44]. The meteorological database of cities used in the study is 4 year average values from 1986 to 2000 as shown in Tables 2 and 3.

3.2. Input variables selection using WEKA

The input variables selection is the first step for developing the ANN model. The input training data: temperature, minimum temperature, maximum temperature, altitude, sunshine hours, latitude and longitude are selected for solar radiation prediction models from Table 2. In the variable selection process, the most relevant input variables for solar radiation prediction have to be evaluated. WEKA is developed by New Zealand government in 1993. It is useful in data mining, business and machine learning. J48 algorithm (a WEKA implementation of c4.5 algorithm) is widely used to construct Decision Tree [45,46]. A decision tree is used for classification rule and represents tree based knowledge. The relevant variable selection for solar radiation prediction has been carried out by using the Decision Tree method. A standard decision tree induced with c4.5 (or possibly ID3 or c5.0) consists of a number of branches, 1 root, some nodes and some leaves. One branch is a chain of nodes from root to a leaf, and each node involves one variable. The occurrence of a variable in a tree provides the information about the importance of associated variable.

To demonstrate the WEKA implementation for relevant variable selection from a input vector X [temperature, minT, maxT, altitude, sunshine hour, latitude, longitude]$_{26\times7}$, we go to the WEKA Explorer using 26 data samples of Table 1. For the selection

of relevant input variable, we choose attribute evaluator and search method and ranks of all variables are observed. The rank of each input variable as determined by WEKA for solar radiation prediction is given in Table 4. The variables longitude and latitude are found to have lowest rank for each month. Therefore for selecting relevant input variables for solar radiation prediction accuracy, longitude and latitude have been omitted from the input vector X and prediction accuracy is to be calculated using ANN based on relevant input variable.

Thus, when the problem of variables selection is complete, we reduce the training data to include only five significant input variables: temperature; min temperature; max temperature; altitude; and sunshine hour. For verifying the authentication of WEKA three ANN models (ANN-1, ANN-2 and ANN-3) are developed to find out prediction accuracy. The ANN-1 model incorporates T, $T_{\min}$, $T_{\max}$, H, SH, Lat and Long. The ANN-2 model uses most significant variables given by WEKA (T, $T_{\min}$, $T_{\max}$, H, SH). The ANN-3 model utilized T, $T_{\min}$, $T_{\max}$ and H for locations where no sun shine duration measuring instruments are installed.

Table 2
Meteorological data and geographical coordinates of 26 Indian cities.

S.No.	City	Lat (°)	Long (°)	H (m)	T (°C)	$T_{\max}$ (°C)	$T_{\min}$ (°C)	SH (hours)
1	Srinagar	34.08	74.79	1730	5.3	15.8	−6.7	6.23
2	New Delhi	28.35	77.12	216	23.8	31.7	13.4	7.74
3	Jodhpur	26.18	73.01	224	25.1	31.2	15.8	8.84
4	Jaipur	26.92	75.82	431	24.6	31.5	15.1	8.05
5	Varanasi	25.45	82.85	81	25.2	31.4	16.8	8.02
6	Patna	25.61	85.13	53	24.7	28.7	17.2	8.33
7	Shillong	25.34	91.53	1598	22.5	26	16.1	5.49
8	Bhopal	23.25	77.42	523	26	34	18.6	8.1
9	Ranchi	23.35	85.33	654	24.3	29.1	17.3	7.92
10	Bhavnagar	21.77	72.15	24	28.3	32.2	24.2	8.46
11	Nagpur	21.09	79.07	311	26.5	34.5	20	7.83
12	Mumbai	19.07	72.51	14	26.7	28.1	24.7	7.73
13	Pune	18.52	73.84	560	25.2	27.5	23.8	7.73
14	Hyderabad	17.36	78.46	536	27	31.7	23.2	7.85
15	Vishakapatnam	17.43	83.14	3	26.6	28.7	23.8	7.86
16	Panjim	15.49	73.81	7	26.6	27.6	25.3	7.78
17	Chennai	13.081	80.27	6	27.7	30.1	25.1	7.76
18	Port Blair	11.61	92.72	73	27.3	28.1	26.4	7.74
19	Minicoy	8.28	73.03	2	27.2	28.1	26.7	7.68
20	Thiruvanathpuram	8.5	76.9	64	26.9	27.6	26.2	7.87
21	Dehradun	30.19	78.02	683	11.3	18.6	1.7	7.85
22	Lucknow	26.45	80.53	128	25.1	32.1	15.8	7.84
23	Hamirpur	31.68	76.52	785	15.9	24	6	7.67
24	Ahmedabad	23.04	72.38	169	27.4	32.2	24.2	7.78
25	Bangalore	12.57	77.38	897	24.7	27.5	21.9	7.79
26	Kolkatta	22.39	88.27	6	25.7	28.4	20.2	7.78

3.3. Solar radiation prediction models with selected inputs

The ANN models (ANN-1, ANN-2 and ANN-3) are developed using artificial neural network fitting tool (nftool), used for prediction. The nftool consists of a standard two layers feed forward neural network trained with Levenberg–Marquardt (LM) algorithm and is suitable for static fitting problems. The training is automatically done with scaled conjugate gradient. The input and target data automatically mapped in the range from −1 to 1 and 60%, 20%, 20% of randomly divided data are used in training, testing and validation respectively. The training data is used to train ANN models with LM algorithm. The testing data have no effect on training and provide an independent measure of network performance during and after training. The validation data are used for measuring generalization capability of network and stop training when generalization comes to an end.

The performance is a plot of mean square error (MSE) with respect to number of epochs. The epochs are one complete sweep of training, testing and validation. The performance plot indicates MSE in training, testing and validation data. The MSE plot in training data has lower curve and has upper curve in validation data set. The network with minimum MSE in validation is called as the trained ANN model. The training automatically stops when validation error stops improving as indicated by an increase in MSE of validation data samples. Training multiple times will generate different results due to random initialization of connection weights and different initial conditions. The stepwise methods to implement

Table 3
Monthly average daily solar radiation data (kWh/m^2/day).

S.No	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Annual
1	1.32	2.71	3.95	5.06	5.62	6.18	5.60	5.20	5.06	3.85	2.56	1.94	4.09
2	3.70	4.56	5.73	6.68	6.78	6.26	5.29	4.94	5.25	4.66	3.92	3.31	5.09
3	4.31	5.05	6.04	6.73	6.97	6.55	5.46	5.41	5.85	5.30	4.49	4.12	5.52
4	4.25	5.01	6.11	7.08	7.25	6.65	5.13	4.88	5.45	5.04	4.27	3.74	5.40
5	3.58	4.76	5.81	6.42	6.39	5.79	4.35	4.80	4.54	4.76	4.01	3.37	4.88
6	3.61	4.72	5.81	6.35	6.29	5.63	4.36	4.64	4.55	4.64	4.08	3.29	4.83
7	3.91	4.63	5.35	5.86	5.11	4.56	4.46	4.14	3.89	4.21	4.34	4.00	4.54
8	4.38	5.20	6.23	7.03	6.75	5.53	4.00	3.80	5.20	5.32	4.72	4.57	5.23
9	4.34	4.91	5.78	6.16	5.88	4.65	4.02	3.85	4.13	4.37	4.26	4.07	4.70
10	4.97	5.81	6.71	7.28	7.37	6.19	4.52	4.48	5.53	5.85	5.09	4.59	5.70
11	4.48	5.33	6.09	6.65	6.55	5.23	4.11	4.10	4.87	5.18	4.54	4.27	5.12
12	4.60	5.41	6.17	6.61	6.48	4.85	3.73	4.03	4.54	5.00	4.61	4.29	5.03
13	4.80	5.71	6.41	6.80	6.99	5.36	4.47	4.35	5.20	5.34	4.90	4.57	5.41
14	5.45	6.11	6.72	6.90	6.63	5.59	5.13	4.87	5.49	5.18	5.01	4.98	5.67
15	4.83	5.55	6.06	6.38	6.16	4.85	4.45	4.54	4.73	4.89	4.55	4.53	5.13
16	5.52	6.22	6.54	6.72	6.56	4.63	4.10	4.40	5.38	5.42	5.32	5.16	5.50
17	4.89	5.85	6.51	6.60	6.26	5.71	5.27	5.20	5.39	4.55	3.99	4.15	5.36
18	5.12	5.85	5.89	5.76	4.37	3.87	3.82	4.02	4.30	4.48	4.65	4.74	4.74
19	4.93	5.61	6.05	5.93	5.05	4.44	4.58	4.88	5.09	5.00	4.63	4.60	5.07
20	5.53	6.12	6.50	5.93	5.44	4.82	4.95	5.27	5.70	5.04	4.60	5.01	5.41
21	3.58	4.40	5.47	6.35	6.95	6.06	5.25	4.80	5.32	5.13	4.22	3.53	5.09
22	4.44	5.43	5.97	6.76	7.14	6.06	5.49	5.32	5.52	5.63	4.78	4.19	5.56
23	2.43	2.87	4.79	5.22	6.14	4.95	4.06	3.48	3.98	4.21	3.16	2.85	4.01
24	4.53	5.43	6.34	6.95	6.99	6.01	4.31	4.30	5.17	5.25	4.65	4.23	5.35
25	5.67	6.48	6.58	6.56	6.35	4.92	4.64	4.48	5.24	5.11	4.84	4.81	5.47
26	3.75	4.35	5.27	5.85	5.73	4.76	4.19	4.32	4.13	4.24	3.84	3.52	4.50

Table 4
Rank of Input Variables by Weka Algorithm for solar radiation prediction.

Month	Temp.	Min. T	Max.T	H	SH	Lat.	Long.
Jan	0.1522	0.1395	0.1089	0.1073	0.0738	0.0639	−0.0248
Feb	0.1523	0.142	0.1122	0.1008	0.0722	0.0639	−0.0251
March	0.1711	0.1408	0.1405	0.1384	0.1017	0.0479	0.0116
April	0.1031	0.0951	0.0909	0.0787	0.0643	0.0441	0.0404
May	0.01146	0.08668	0.046	0.05148	0.0486	0.00738	0.0019
June	0.01615	0.04742	0.00352	0.0011	0.02052	−0.00164	−0.01117
July	0.00657	0.01605	0.0125	0.01863	0.00272	−0.01248	−0.02144
Aug.	0.015499	0.000614	0.000859	0.009573	−0.00408	−0.010689	−0.010926
Sep.	−0.01775	0.10651	−0.00954	0.01511	0.03343	−0.03026	−0.02267
Oct.	0.061942	0.038207	0.053018	0.075084	0.067926	0.000164	0.035788
Nov.	0.1168	0.0977	0.0867	0.0765	0.0546	0.0343	−0.0175
Dec.	0.1102	0.1026	0.0762	0.0691	0.0498	0.044	−0.0219

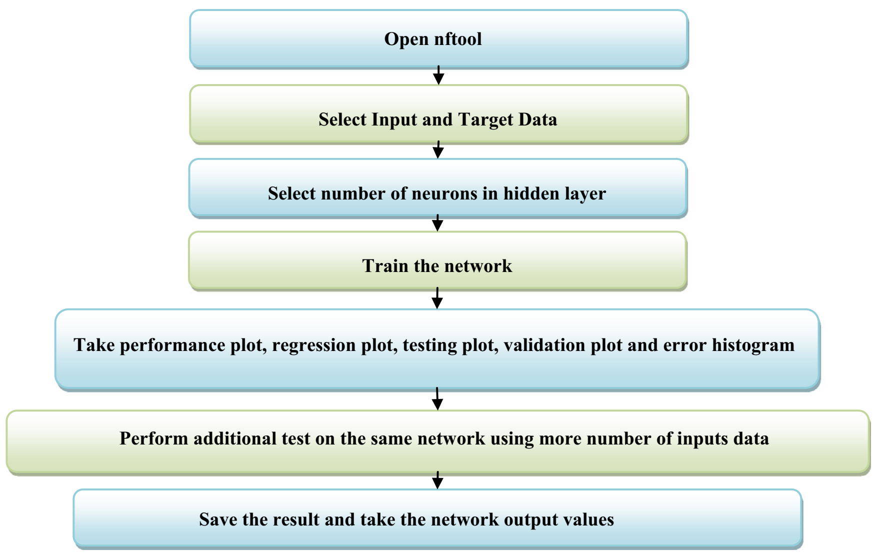

Fig. 2. Implementation of neural network fitting tool (nftool).

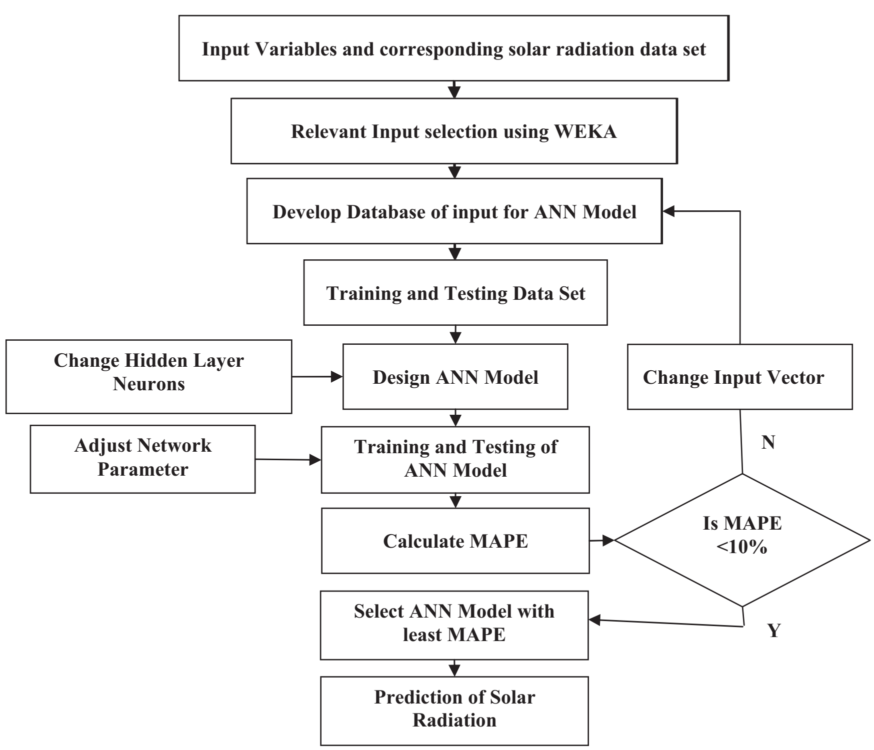

Fig. 3. Proposed algorithm for solar radiation prediction.

Table 5
Statistical Error Evaluation in ANN-1 model.

Sensitivity test of hidden layer neurons	MLP structure	R for training	Maximum MAPE for testing	Selection of ANN architecture
The number of inputs (I_n) is 7, number of outputs (O_n) is 12 (monthly average solar radiation) and number of samples (S_n) is 22. Therefore the hidden layer neurons (H_n) varies from 9 to 19.	**7-9-1**	**92.45**	**20.12**	The ANN architecture (7-9-1) with 7 neurons in input, 9 neurons in hidden layer and 1 neurons in output layer is best as it has least MAPE
	7-10-1	92.96	34.93	
	7-11-1	93.17	37.83	
	7-12-1	92.52	28.17	
	7-13-1	93.21	23.81	
	7-14-1	93.68	38.14	
	7-15-1	90.08	28.33	
	7-16-1	91.30	20.53	
	7-17-1	94.48	31.68	
	7-18-1	92.42	27.97	
	7-19-1	93.85	25.04	

Table 6
Statistical error evaluation in the ANN-2 model.

Sensitivity test of hidden layer neurons	MLP structure	R for training	Maximum MAPE for testing	Selection of ANN architecture
The number of inputs (I_n) is 5, number of outputs (O_n) is 12 (monthly average solar radiation) and number of samples (S_n) is 22. Therefore the hidden layer neurons (H_n) varies from 8 to 18.	5-8-1	91.43	15.96	The ANN architecture (5-10-1) with 5 neurons in input, 10 neurons in hidden layer and 1 neurons in output layer is best as it has least MAPE
	5-9-1	86.77	10.90	
	5-10-1	**91.22**	**6.89**	
	5-11-1	84.97	15.67	
	5-12-1	92.79	10.42	
	5-13-1	93.066	14.71	
	5-14-1	91.61	14.21	
	5-15-1	88.75	16.96	
	5-16-1	91.21	16.88	
	5-17-1	90.38	23.24	
	5-18-1	92.55	15.17	

Table 7
Statistical error evaluation in the ANN-3 model.

Sensitivity test of hidden layer neurons	MLP structure	R for training	Maximum MAPE for testing	Selection of ANN architecture
The number of inputs (I_n) is 4, number of outputs (O_n) is 12 (monthly average solar radiation) and number of samples (S_n) is 22. Therefore the hidden layer neurons (H_n) varies from 7 to 18.	4-7-1	89.48	18.06	The ANN architecture (4-10-1) with 4 neurons in input, 10 neurons in hidden layer and 1 neurons in output layer is best as it has least MAPE
	4-8-1	92.62	10.84	
	4-9-1	91.34	13.76	
	4-10-1	**93.69**	**9.04**	
	4-11-1	82.51	14.14	
	4-12-1	88.63	19.76	
	4-13-1	94.26	18.40	
	4-14-1	92.50	13.64	
	4-15-1	90.26	21.51	
	4-16-1	91.88	16.66	
	4-17-1	92.59	28.52	
	4-18-1	94.27	44.52	

nftool are shown in Fig. 2, Matlab code in Appendix A and proposed algorithm in Fig. 3.

The number of neurons in hidden layer is evaluated by Eq. (1) [47,48], where H_n and S_n are number of hidden layer neurons and number of data samples used in the ANN model, I_n and O_n denotes number of input and output parameters.

$$H_n = \frac{I_n + O_n}{2} + \sqrt{S_n} \tag{1}$$

The sensitivity test is performed to validate the number of hidden layer neurons by calculating change in prediction error (MAPE) when number of hidden layer neurons is changed ± 5 from hidden layer neurons calculated by Eq. (1). The sensitivity analysis of hidden layer neurons for ANN models are don are shown in Tables 5–7. The MAPE is given by Eq. (2) and ANN architecture with least MAPE is used for prediction of solar radiation.

$$\mathrm{MAPE} = \left(\frac{1}{n} \sum_{i=1}^{n} \left| \frac{SR_{i(ANN)} - SR_{i(actual)}}{SR_{i(actual)}} \right| \right) \times 100 \tag{2}$$

4. Results and discussion

The prediction accuracy is evaluated with MAPE given by Lewis [49]. The MAPE $\leq 10\%$ indicates high prediction accuracy, $10\% \leq$ MAPE $\leq 20\%$ indicates good prediction, $20\% \leq \mathrm{MAPE} \leq 50\%$ indicates reasonable prediction, $\mathrm{MAPE} \geq 50\%$ indicates inaccurate forecasting. The maximum MAPE of testing cities for ANN-1, ANN-2 and ANN-3 models are found to be 20.12%, 6.89% and 9.04% respectively, showing that after removing less influencing

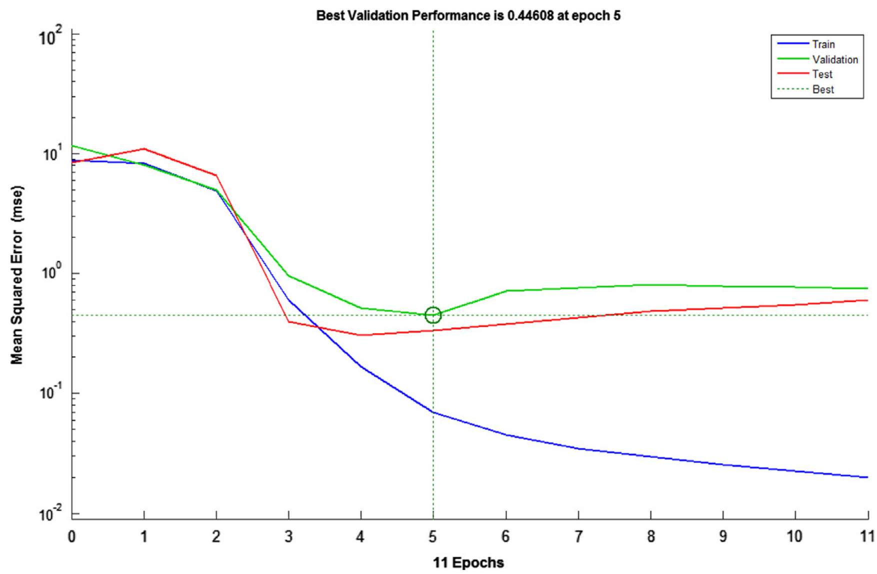

Fig. 4. Performance plot of the ANN-2 model during training.

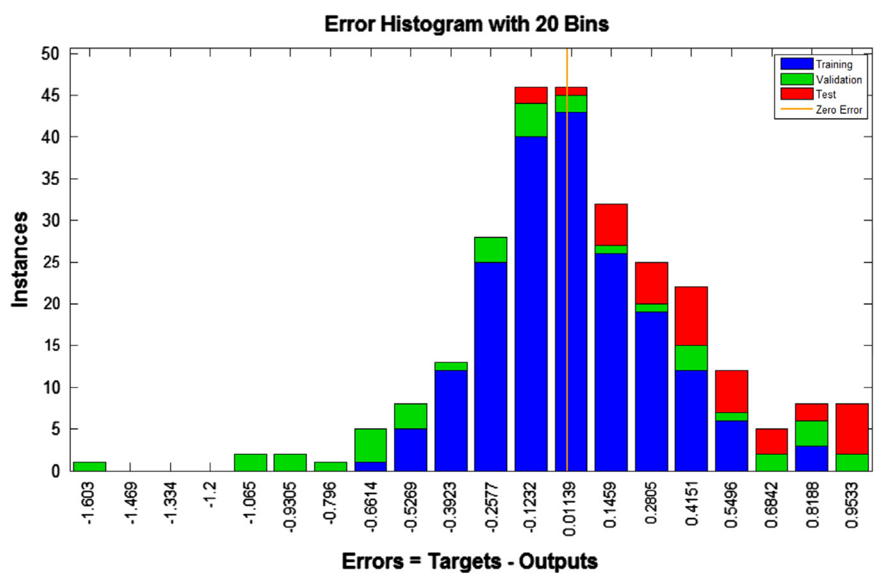

Fig. 5. Error histogram plot of the ANN-2 model. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

parameters (Lat, Long) in the ANN-2 model, the prediction accuracy is increased by 13.23%. Therefore WEKA can be used for identifying relevant input parameters for solar radiation prediction. The MAPE of ANN-3 is more than ANN-2, showing sunshine is vital parameter for solar radiation prediction but it can be used for prediction where sunshine hour measured data are not available.

The performance plot of the ANN-2 model demonstrates that mean square error becomes minimum as the number of epochs is increasing (Fig. 4). The test set error and validation set error has comparable characteristics and no major over fitting has happened near epoch 5 (where best validation performance has taken place).

The error histogram plot is shown in Fig. 5 to present further authentication of network performance. It points towards outliers, which are data features where the fit is drastically not as good as than the best part of data. The blue, green and red bars signify training data, validation data and testing data respectively. The largest part of data coincides with zero error line which offers a scheme to verify the outliers to decide if the data is imperfect, or if those data features are unlike than the leftover of data set.

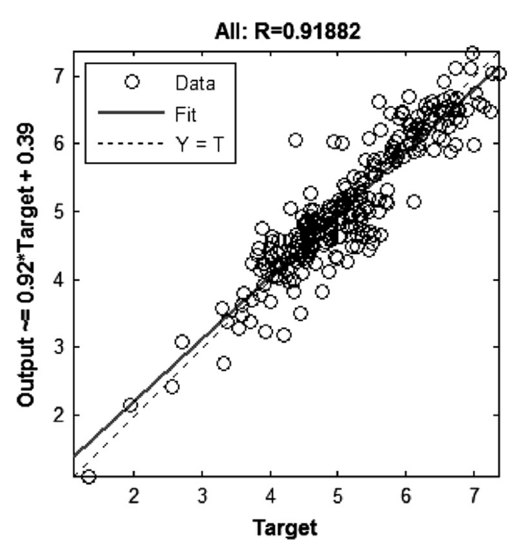

Fig. 6. The ANN-2 model regression plot.

Table 8
Predicted solar radiation data (kWh/m^2/day) of 15 towns of Himachal Pradesh.

Town	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Annual
Chamba	3.710	4.464	5.45	6.065	6.715	6.516	5.518	5.267	5.651	5.123	4.200	3.827	5.209
Kangra	**4.032**	**4.721**	**5.67**	**6.237**	**6.913**	**6.565**	**5.570**	**5.285**	**5.771**	**5.318**	**4.435**	**4.072**	**5.383**
Hamirpur	3.122	4.039	5.277	6.632	7.025	5.501	4.868	4.069	4.935	4.882	3.964	3.02	4.778
Bilaspur	2.827	3.81	5.118	6.722	7.13	5.392	4.94	3.983	4.797	4.782	3.767	2.566	4.653
Shimla	**0.563**	**2.339**	**3.966**	**4.706**	**5.689**	**5.672**	**4.749**	**4.376**	**3.635**	**3.39**	**2.53**	**1.516**	**3.594**
Una	2.686	3.7	5.043	6.766	7.18	5.34	4.978	3.943	4.734	4.733	3.671	2.349	4.594
Mandi	3.824	4.6	5.675	6.422	6.767	5.77	4.729	4.294	5.295	5.102	4.439	4.124	5.087
Solan	4.399	5.191	6.171	6.177	6.324	5.671	4.205	4.128	5.458	5.116	4.975	5.437	5.271
Kullu	2.976	3.902	4.96	5.71	6.273	6.215	5.303	5.097	5.297	4.66	3.69	3.2543	4.778
Nahan	3.227	4.148	5.377	6.524	6.727	5.123	4.329	3.69	4.742	4.749	4.093	3.419	4.679
Nalagarh	3.773	4.509	5.745	7.02	7.243	5.914	5.672	4.447	5.929	5.024	4.234	3.624	5.261
Kaza	2.883	2.194	4.369	4.838	7.119	6.531	5.489	4.578	2.691	3.543	2.743	0.568	3.962
Keylong	2.941	2.097	4.272	4.811	7.038	6.385	5.342	4.456	2.575	3.483	2.664	0.4643	3.877
Kalpa	2.725	2.129	4.254	4.788	6.895	6.324	5.286	4.449	2.688	3.468	2.659	0.598	3.855
Reckong Peo	1.577	2.022	3.849	4.756	6.331	6.372	5.688	4.955	3.587	3.426	2.285	0.613	3.788

Fig. 7. Predicted annual average global solar radiation of Himachal Pradesh, India (kWh/m^2/day).

The correlation coefficient (R-value) determines the association among outputs and targets value of the ANN-2 model. R value of 1and 0 measures a strong, random association respectively. The perfect fit indicates that the data should fall along 45° line (slope is close to 1), means network output is equal to targets. The R-value is 0.91 and slope is 0.92 is achieved during whole dataset; proving that the ANN-2 model (nftool) predicts solar radiation close to measured value (Fig. 6).

4.1. Solar radiation predicted by the ANN-3 model

The estimation of solar energy resource potential the western Himalayan state of Himachal Pradesh India is another objective of our study. However, the sunshine hour data are not measured at most of the meteorological stations in the state. Therefore the ANN-3 model which excludes sunshine hour as input with lesser accuracy than the ANN-3 model is used for the prediction of solar radiation of 15 towns of the state as shown in Table 8.

The annual global solar radiation in Himachal Pradesh is found to vary from 3.59 to 5.38 kWh/m^2/day (Fig. 7), which is in the range of values given in India solar resource map (by NREL and Solar Energy Centre) [50]. Therefore the predicted solar radiation of H.P. towns is accurate enough to be used for various solar energy applications.

5. Conclusion

The present work has shown the powerful nature of the WEKA to evaluate the most influencing input parameters in prediction of

solar radiation using ANN. The most relevant input variables for predicting the solar radiation are found to be temperature, maximum temperature, minimum temperature, altitude above mean sea level and sunshine hours. It is found that latitude and longitude have minimum effect on solar radiation prediction. The maximum MAPE for ANN-1, ANN-2 and ANN-3 models are 20.12%, 6.89% and 9.04% respectively, showing high accuracy of ANN-2 which utilizes most relevant input variables. The developed ANN-2 model can be used for prediction of solar radiation at any sites in India. Therefore it can be used for assessment of solar energy resource potential. The predicted solar radiation using the ANN-3 model for the cities of Himachal Pradesh state varies from 3.59 to 5.38 kWh/m^2/day yearly, showing a good solar potential, which can be utilized for installation of solar photovoltaic power plants, solar hybrid systems and solar thermal applications. Additionally the state has 14% barren or uncultivable land with south facing mountain slopes which can also be utilized for solar power generation.

Further studies to estimate the solar potential of the region with greater accuracy can be undertaken. Future research is to be focused on to find most relevant input parameters from other meteorological variables with improved prediction accuracy of different ANN models.

Appendix A. MATLAB code for solar radiation using nftool

```
clc
clear all
close all
% ip-input data (geographical and meteorological variable)
% tr-target data (solar radiation)
% p-testing data (geographical and meteorological variable).
ip=xlsread('inputdata.xlsx'); tr=xlsread('targetdata.xlsx');
  p=xlsread('testingdata.xlsx'); inputs=ip′; targets =tr;
% Create a Fitting Network
hiddenLayerSize=10; % Select according to problem
net=fitnet(hiddenLayerSize)
% Train Parameters
net.divideParam.trainRatio=70/100; % Select according to
  problem
net.divideParam.valRatio=15/100; % Select according to
  problem
net.divideParam.testRatio=15/100; % Select according to
  problem
% Train the Network
[net,tr]=train(net,inputs,targets);
% Test the Network
outputs=net(inputs); errors=gsubtract(targets,outputs);
  performance=perform(net,targets,outputs).
% View the Network
view (net);
a=sim(net,p′); % output of testing input data
a1=sim(net,ip′); % output of training input data
% weight and bias to hidden layer
b1=net.b{1,1}; % bias to hidden layer
W1=net.IW{1,1}; % weight to hidden layer
% weight and bias to output layer
b2=net.b{2,1}; % bias to output layer
W2=net.LW{2,1}; % weight to output layer
% Plots
plotperform(tr); plottrainstate(tr); plotfit(net,inputs,targets);
  plotregression(targets,outputs); ploterrhist(errors)
% Statistical Results
SRi(ANN);% solar radiation predicted by neural network
SRi(actual); % measured values of solar radiation
MAPE; % Mean Absolute Percentage Error
MAPE=1/12*sum(abs(SRi(ANN)- SRi(actual)/ SRi(actual)))*100
R2; Absolute fraction of variance
R2=(1−(sum(SRi(ANN)−SRi(actual))\widehat2)/sum
  (SRi(actual)))
```

References

[1] Bakirci K. Models of solar radiation with hours of bright sunshine: a review. Renew Sustain Energy Rev 2009;13:2580–8.

[2] Khatib T, Mohamed A, Sopian K. A review of solar energy modeling techniques. Renew Sustain Energy Rev 2012;16:2864–9.

[3] Şenkal O, Kuleli T. Estimation of solar radiation over Turkey using artificial neural network and satellite data. Appl Energy 2009;86:1222–8.

[4] Mubiru J, Banda EJKB. Estimation of monthly average daily global solar irradiation using artificial neural networks. Sol Energy 2008;82:181–7.

[5] Jiang Y. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. Energy Policy 2008;36:3833–7.

[6] Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C. Comparative study of Ångström′s and artificial neural networks′ methodologies in estimating global solar radiation. Sol Energy 2005;78:752–62.

[7] Yadav AK, Chandel SS. Solar radiation prediction using artificial neural network techniques: a review. Renew Sustain Energy Rev 2013. http://dx.doi.org/10.1016/j.rser.2013.08.055i.

[8] Ramachandra TV, Jain R, Krishnadas G. Hotspots of solar potential in India. Renew Sustain Energy Rev 2011;15:3178–86.

[9] G Krishnadas, TV Ramachandra, Scope for renewable energy in Himachal Pradesh, India –a study of solar and wind resource potential. Lake 2010: Wetlands, Biodiversity and Climate Change. p. 1–10.

[10] 〈https://eosweb.larc.nasa.gov/cgi-bin/sse/print.cgi?accuracy.txt〉.

[11] I. Karakoti, A Datta, SK Singh. Validation of satellite based solar radiation data with ground measurements. In: Proceedings of the 14th annual international conference and exhibition on geospatial information technology and application; 2012. p. 1–10.

[12] AI-Alawi SM, AI-Hinai HA. An ANN based approach for predicting global radiation in locations with no direct measurement instrumentation. Renew Energy 1998;14(1–4):199–204.

[13] Sözen A, Arcaklioğlu E, Özalp M, Kanit EG. Use of artificial neural networks for mapping of solar potential in Turkey. Appl Energy 2004;77:273–86.

[14] Sözen A, Arcaklioğlu E, Özalp M. Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data. Energy Convers Manag 2004;45:3033–52.

[15] Mohandes M, Rehman S, Halawani TO. Estimation of global solar radiation using artificial neural networks. Renew Energy 1998;14(1–4):179–84.

[16] Ouammi A, Zejli D, Dagdougui H, Benchrifa R. Artificial neural network analysis of Moroccan solar potential. Renew Sustain Energy Rev 2012;16: 4876–89.

[17] Rehman S, Mohandes M. Estimation of diffuse fraction of global solar radiation using artificial neural networks. Energy Sources, Part A 2009;31: 974–84.

[18] Azeez MAA. Artificial neural network estimation of global solar radiation using meteorological parameters in Gusau, Nigeria. Arch Appl Sci Res 2011;3 (2):586–95.

[19] Linares-Rodriguez A, Ruiz-Arias JA, Pozo-Vazquez D, Tovar-Pescador J. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 2013;61:636–45.

[20] Kisi O. Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. Energy 2014;64:429–36.

[21] Mostafavi ES, Ramiyani SS, Sarvar R, Moud HI, Mousavi SM. A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran. Energy 2013;49:204–10.

[22] Linares-Rodríguez A, Ruiz-Arias JA, Pozo-Vázquez D, Tovar-Pescador J. Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. Energy 2011;36:5356–65.

[23] Koca A, Oztop HF, Varol Y, Koca GO. Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. Expert Syst Appl 2011;38:8756–62.

[24] Khatib T, Mohamed A, Sopian K, Mahmoud M. Solar energy prediction for Malaysia using artificial neural networks. Int J Photoenergy 2012:1–16.

[25] Khatib T, Mohamed A, Mahmoud M, Sopian K. Modeling of daily solar energy on a horizontal surface for five main sites in Malaysia. Int J Green Energy 2011;8:795–819.

[26] Elminir HK, Areed FF, Elsayed TS. Estimation of solar radiation components incident on Helwan site using neural networks. Sol Energy 2005;79:270–9.

[27] Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C. Comparative study of Ångström′s and artificial neural networks′ methodologies in estimating global solar radiation. Sol Energy 2005;78:752–62.

[28] Alam S, Kaushik SC, Garg SN. Computation of beam solar radiation at normal incidence using artificial neural network. Renew Energy 2006;31:1483–91.

[29] Jiang Y. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. Energy Policy 2008;36:3833–7.
[30] Mubiru J, Banda EJKB. Estimation of monthly average daily global solar irradiation using artificial neural networks. Sol Energy 2008;82:181–7.
[31] Şenkal O, Kuleli T. Estimation of solar radiation over Turkey using artificial neural network and satellite data. Appl Energy 2009;86:1222–8.
[32] Jiang Y. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. Energy 2009;34:1276–83.
[33] Benghanem M, Mellit A, Alamri SN. ANN-based modeling and estimation of daily global solar radiation data: a case study. Energy Convers Manag 2009;50:1644–55.
[34] Fadare DA. Modelling of solar energy potential in Nigeria using an artificial neural network model. Appl Energy 2009;86:1410–22.
[35] Azadeh A, Maghsoudi A, Sohrabkhani S. An integrated artificial neural networks approach for predicting global radiation. Energy Convers Manag 2009;50:1497–505.
[36] Sözen A, Arcaklioğlu E. Solar potential in Turkey. Appl Energy 2005;80:35–45.
[37] Rahimikhoob A. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. Renew Energy 2010;35:2131–5.
[38] Hargreaves GH, Samani ZA. Reference crop evapotranspiration from temperature. Appl Engine Agric 1985;1(2):96–9.
[39] Hasni A, Sehli A, Draoui B, Bassou A, Amieur B. Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. Energy Procedia 2012;18:531–7.
[40] Yildiz BY, Şahin M, Şenkal O, Pestemalci V, Emrahoğlu N A. Comparison of two solar radiation models using artificial neural networks and remote sensing in Turkey. Energy Sources Part A 2013;35:209–17.
[41] Rumbayan M, Abudureyimu A, Nagasaka K. Mapping of solar energy potential in Indonesia using artificial neural network and geographical information system. Renew Sustain Energy Rev 2012;16:1437–49.
[42] ⟨https://eosweb.larc.nasa.gov/sse/RETScreen/⟩. [accessed 01.06.13].
[43] Mani A. Handbook of solar radiation data for India. New Delhi: Allied Publishers Pvt. Ltd; 1981.
[44] ⟨http://indiaenvironmentportal.org.in/files/srd-sec.pdf⟩.
[45] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update; SIGKDD Explorations 2009; 11(1). [accessed 02.02.2013] ⟨http://dx.doi.org/http://www.cs.waikato.ac.nz/~eibe/pubs/weka_update.pdf⟩.
[46] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. A Book of Morgan Kaufmann Publishers. 3rd ed.2011.
[47] Chow SKH, Lee EWM, Li DHW. Short-term prediction of photovoltaic energy generation by intelligent approach. Energy Build 2012;55:660–7.
[48] Frederick M. Neuroshell 2 Manual, Ward Systems Group Inc., 1996.
[49] Lewis CD. International and business forecasting methods. London: Butterworths; 1982.
[50] ⟨http://mnre.gov.in/sec/GHI_Annual.jpg⟩.